Nonlinear approximation with dictionaries. II. Inverse estimates

Rémi Gribonval, Morten Nielsen

NONLINEAR APPROXIMATION WITH DICTIONARIES. II. INVERSE ESTIMATES.

RÉMI GRIBONVAL AND MORTEN NIELSEN

ABSTRACT. In this paper, which is the sequel to [GN04a], we study inverse estimates of the Bernstein type for nonlinear approximation with structured redundant dictionaries in a Banach space. The main results are for *blockwise incoherent dictionaries* in Hilbert spaces, which generalize the notion of *joint block-diagonal mutually incoherent bases* introduced by Donoho and Huo. The Bernstein inequality obtained for such dictionaries is proved to be sharp, but it has an exponent that does not match that of the corresponding Jackson inequality.

1. INTRODUCTION

Let X be a separable Banach space, and $\mathcal{D} = \{g_k, k \geq 1\}$ a countable family of unit vectors, $\|g_k\|_X = 1$, which will be called a **dictionary** whenever it spans a dense subspace of X. Our main purpose in this paper is to study the **approximation spaces** $\mathcal{A}_q^\alpha(\mathcal{D}, X)$ associated with best m-term approximation. Let us recall the definition of $\mathcal{A}_q^\alpha(\mathcal{D}, X)$. The (nonlinear) set of all linear combinations of at most m elements from $\mathcal{D}$ is

$$\Sigma_m(\mathcal{D}) := \left\{ \sum_{k \in I_m} c_k g_k, I_m \subset \mathbb{N}, \mathrm{card}(I_m) \leq m, c_k \in \mathbb{C} \right\}.$$

For any given $f \in X$, the error associated to the *best m-term* approximation to f from $\mathcal{D}$ is given by

$$\sigma_m(f, \mathcal{D})_X := \inf_{h \in \Sigma_m(\mathcal{D})} \|f - h\|_X.$$

The *best m-term approximation spaces* are defined as :

$$\mathcal{A}_q^\alpha(\mathcal{D}, X) := \left\{ f \in X, \|f\|_{\mathcal{A}_q^\alpha(\mathcal{D},X)} := \|f\|_X + |f|_{\mathcal{A}_q^\alpha(\mathcal{D},X)} < \infty \right\}$$

where $|\cdot|_{\mathcal{A}_q^\alpha(\mathcal{D},X)} := \|\{\sigma_m(f, \mathcal{D})_X\}_{m \geq 1}\|_{\ell_q^{1/\alpha}}$ is defined using the Lorentz (quasi)norm, see e.g. [DL93].

Stechkin, DeVore and Temlyakov have derived the following nice characterization when the dictionary is an orthonormal basis in a Hilbert space.

Theorem 1.1 ([Ste55, DT96]). *If $\mathcal{B}$ is an orthonormal basis in a Hilbert space $\mathcal{H}$ then, for $0 < \tau = (\alpha + 1/2)^{-1} < 2$ and $0 < q \leq \infty$,*

$$\mathcal{A}_q^\alpha(\mathcal{B}, \mathcal{H}) = \mathcal{K}_q^\tau(\mathcal{B}, \mathcal{H})$$

Key words and phrases. Nonlinear approximation, Bernstein inequality, mutually incoherent bases.

with equivalent (quasi)norms, where

$$\mathcal{K}_q^\tau(\mathcal{B},\mathcal{H}) := \left\{ f \in \mathcal{H}, |f|_{\mathcal{K}_q^\tau(\mathcal{B},\mathcal{H})} := \|\{\langle f, g_k\rangle\}_{k\geq 1}\|_{\ell_q^\tau} < \infty \right\}.$$

In recent papers, similar results were obtained whenever $\mathcal{B}$ is an *almost-greedy basis* in a general Banach space [GN01, DKKT01, KP01], or a redundant system of *framelets* in $L^p(\mathbb{R})$ [GN04c]. The goal of this paper is to generalize Theorem 1.1 to some redundant dictionaries. Based on examples in [GN01, GN04b] we know that we need to require some structure of $\mathcal{D}$.

1.1. **Direct estimates.** In the prequel [GN04a] to the present paper, we showed that the ℓ_1^p-*hilbertian property* of $\mathcal{D}$, with $p > 1$, is (almost) equivalent to the Jackson embedding $\mathcal{K}_q^\tau(\mathcal{D}, X) \hookrightarrow \mathcal{A}_q^\alpha(\mathcal{D}, X)$ with $0 < \tau = (\alpha + 1/p)^{-1} < p$, where $\mathcal{K}_q^\tau(\mathcal{D}, X)$ is a sparsity space introduced by DeVore and Temlyakov [DT96].

Definition 1.2. A dictionary $\mathcal{D}$ is ℓ_q^τ-hilbertian if the operator $T : \{c_k\} \mapsto \sum_k c_k g_k$ defined on the space ℓ^0 of finite sequences $\mathbf{c} = \{c_k\}$ extends to a continuous operator from ℓ_q^τ to X.

Moreover, we proved that if $\mathcal{D}$ is ℓ_1^p-hilbertian, $p > 1$, we have the representation $\mathcal{K}_q^\tau(\mathcal{D}, X) = T\ell_q^\tau$ of the sparsity spaces for $0 < \tau < p$ with norm

$$|f|_{\mathcal{K}_q^\tau(\mathcal{D})} = \min\{\|\mathbf{c}\|_{\ell_q^\tau} : f = T\mathbf{c}\}. \tag{1.1}$$

The latter representation is quite a bit simpler to handle than the general definition of a sparsity space for an arbitrary dictionary, which can be found in [DT96, GN04a].

1.2. **Inverse estimates.** In this paper, we focus our attention on getting inverse embeddings of the Bernstein type

$$\mathcal{A}_q^\alpha(\mathcal{D}, X) \hookrightarrow \mathcal{K}_q^\tau(\mathcal{D}, X) \tag{1.2}$$

General results of approximation theory [DL93, Chap. 7] relate the embedding (1.2) to a Bernstein inequality for $\mathcal{K}_q^\tau(\mathcal{D})$ with exponent α :

$$|f_m|_{\mathcal{K}_q^\tau(\mathcal{D})} \leq C m^\alpha \|f\|_X, \qquad m \geq 1, f_m \in \Sigma_m(\mathcal{D}). \tag{1.3}$$

1.3. **Fragility of the inverse estimates.** Based on examples in [GN01] we know that, in addition to the ℓ_1^p-hilbertian assumption, we need to require some structure of $\mathcal{D}$ to get (1.3), and we may have to restrict our ambitions to proving a Bernstein inequality with exponent $\alpha > 1/\tau - 1/p$. Recently, Gröchenig introduced the notion of *localized frames* [Grö03, Grö04] and suggested that, in Hilbert spaces, localization might be a sufficient condition to have a Bernstein inequality. In [GN04b] we disproved this fact using a simple counter-example where the dictionary was a localized union of two bases, and yet no Bernstein inequality could be satisfied for any exponent $0 < \alpha < \infty$. In contrast to this negative result which showed the general fragility of Bernstein inequalities, we prove in this paper that *blockwise incoherent* dictionaries in Hilbert spaces actually satisfy a somewhat robust Bernstein inequality, where the exponent α of the corresponding Jackson inequality is generally not matched but is at most doubled.

The structure of the paper is as follows. In Section 2 we study dictionaries in finite dimensional Hilbert spaces, and obtain our first main result: a Bernstein inequality which is related to the *coherence* of the dictionary. For dictionaries that are the union of two orthonormal bases, the coherence coincide with the notion of *mutual coherence* of the bases introduced by Donoho and Huo [DH01]. We provide an example to show that the exponent of the obtained Bernstein inequality –which does not match that of the corresponding Jackson inequality– is sharp in the sense that it cannot be generally improved.

In Section 3 we consider the class of decomposable dictionaries in (possibly infinite dimensional) Banach spaces and prove that a "global" Bernstein inequality can be obtained by proving "blockwise" Bernstein inequalities. We combine this with the results of the first section to prove a Bernstein inequality for *blockwise incoherent dictionaries* in Hilbert spaces. Using results on *Grassmannian frames*, we show that this Bernstein estimate is valid for dictionaries that can be highly redundant. More classical examples include pairs of *jointly block-diagonal mutually incoherent bases* such as the (Meyer-Lemarié wavelets, real bi-sinusoids) example of Donoho and Huo [DH01] or simply the (Haar,Walsh) dictionary.

In the final part of the paper (Section 4) we apply the results to get two-sided embeddings between the approximation classes and function spaces that measure smoothness in terms of a mixture of Besov spaces and spaces related to the Wiener algebra.

2. Incoherent dictionaries in finite dimension

In this section we will prove a Bernstein inequality for dictionaries in X a finite dimensional Hilbert space. In finite dimension, all norms are equivalent so the important result in this Bernstein inequality will be that the constant C in Eq. (1.3) only depends on the *separation factor* of the dictionary, which needs not depend on the dimension of the Hilbert space.

The notion of separation factor of a dictionary is motivated by the concept of *mutual incoherence* between orthonormal bases which was introduced by Donoho and Huo in [DH01]: consider $\mathcal{B}_1 = \{g_k^1\}_{k=1}^N$ and $\mathcal{B}_2 = \{g_k^2\}_{k=1}^N$ two orthonormal bases of the same Hilbert space $\mathcal{H}$ of dimension N. It is not difficult to check (see [DH01, Lemma VII.2]) that $M(\mathcal{B}_1, \mathcal{B}_2) := \max_{k,l} |\langle g_k^1, g_l^2\rangle| \geq 1/\sqrt{N}$. Two such bases are said to be *most mutually incoherent* if $M(\mathcal{B}_1, \mathcal{B}_2) = 1/\sqrt{N}$. The prime example is the basis pair $(\mathcal{B}_1, \mathcal{B}_2)$=(Spikes,Sinusoids).

For a general dictionary $\mathcal{D}$ in a Hilbert space $\mathcal{H}$ (not necessarily finite dimensional) the **coherence** is defined [GN03, DE03] as

$$M(\mathcal{D}) := \max_{k \neq l} |\langle g_k, g_l\rangle| \,. \tag{2.1}$$

It naturally generalizes the measure of *mutual coherence* $M(\mathcal{B}_1, \mathcal{B}_2)$ to dictionaries that are not necessarily the union of two orthonormal bases.

When $\mathcal{H}$ is of finite dimension N we define the **separation factor**

$$S(\mathcal{D}) := M(\mathcal{D})\sqrt{N}. \tag{2.2}$$

Clearly, if $\mathcal{D}$ is an orthonormal basis, then $S(\mathcal{D}) = 0$, and if one can extract at least one orthonormal basis $\mathcal{B}$ from $\mathcal{D}$, then $M(\mathcal{D}) \geq 1/\sqrt{N}$ and $S(\mathcal{D}) \geq 1$. We will say that $\mathcal{D}$ is **perfectly incoherent** if $S(\mathcal{D}) = 1$. We can now formulate the following Bernstein inequality which is the first main result of this paper.

Theorem 2.1. *Let $\mathcal{D}$ be a dictionary in a finite dimensional Hilbert space $\mathcal{H}$ of dimension N, and assume that $\mathcal{D}$ contains an orthonormal basis $\mathcal{B}$. For any $0 < \tau < 2$, the Bernstein inequality for $\mathcal{K}_\tau^\tau(\mathcal{D})$ holds with exponent $\alpha = 2(1/\tau - 1/2)$:*

$$|f_m|_{\mathcal{K}_\tau^\tau(\mathcal{D})} \leq Cm^\alpha \|f_m\|_{\mathcal{H}}, \qquad m \geq 1,\ f_m \in \Sigma_m(\mathcal{D}),$$

where

$$C \leq \max\left(\sqrt{2}, \left(2S(\mathcal{D})\right)^{2/\tau - 1}\right). \tag{2.3}$$

Moreover, the exponent $\alpha = 2(1/\tau - 1/2)$ is sharp for the class of perfectly incoherent dictionaries.

Note that the sharpness of Theorem 2.1 only means that *among the class of perfectly incoherent dictionaries*, there is a subfamily for which the estimate cannot be improved. It does not rule out that some other family of *particular* perfectly incoherent dictionaries satisfy an estimate with an improved exponent. The proof of Theorem 2.1 will appear at the end of this section. First we need a few lemmas. The first lemma is an almost trivial remark, but it will be quite useful later.

Lemma 2.2. *Let $\mathcal{D}$ be a dictionary in a finite dimensional Hilbert space $\mathcal{H}$ of dimension N, and assume that $\mathcal{D}$ contains an orthonormal basis $\mathcal{B}$. For any $0 < \tau < 2$ we have for all $f \in \mathcal{H}$*

$$|f|_{\mathcal{K}_\tau^\tau(\mathcal{D})} \leq N^{1/\tau - 1/2} \|f\|_{\mathcal{H}}. \tag{2.4}$$

Proof. Because $\mathcal{B} \subset \mathcal{D}$ is an orthonormal system, we have the trivial estimate $|f|_{\mathcal{K}_\tau^\tau(\mathcal{D})} \leq |f|_{\mathcal{K}_\tau^\tau(\mathcal{B})} \leq N^{1/\tau - 1/2} \|f\|_{\mathcal{H}}$. □

Lemma 2.3. *Let $\mathcal{D}$ be a dictionary in a Hilbert space $\mathcal{H}$. For $m \geq 1$ and any $f_m = \sum_{k \in I_m} a_k g_k$ with card$(I_m) \leq m$, we have*

$$(1 + M(\mathcal{D})(1 - m)) \sum_{k \in I_m} |a_k|^2 \leq \|f\|_{\mathcal{H}}^2. \tag{2.5}$$

Proof. We develop the expression $\|f\|_{\mathcal{H}}^2 = \langle f, f \rangle$, using the fact that $|\langle g_k, g_l \rangle| \leq M(\mathcal{D})$, $k \neq l$, to get

$$\begin{aligned}
\|f\|^2 &\geq \sum_{k \in I_m} |a_k|^2 - M(\mathcal{D}) \sum_{k,l \in I_m} |a_k||a_l| + M(\mathcal{D}) \sum_{k \in I_m} |a_k|^2 \\
&\geq (1 + M(\mathcal{D})) \sum_{k \in I_m} |a_k|^2 - M(\mathcal{D}) (\sum_{k \in I_m} |a_k|)^2 \\
&\geq (1 + M(\mathcal{D})) \sum_{k \in I_m} |a_k|^2 - M(\mathcal{D}) \left(m^{1/2} (\sum_{k \in I_m} |a_k|^2)^{1/2}\right)^2 \\
&\geq (1 + M(\mathcal{D})(1 - m)) \sum_{k \in I_m} |a_k|^2.
\end{aligned}$$

□

We combine the lemmas to get the following result.

Corollary 2.4. *Let $\mathcal{D}$ be a dictionary in a Hilbert space $\mathcal{H}$. For any $0 < \tau < 2$, $\lambda > 1$, $m \le 1 + (\lambda M(\mathcal{D}))^{-1}$ and any $f_m \in \Sigma_m(\mathcal{D})$ we have the inequality*

$$|f_m|_{\mathcal{K}_\tau^\tau(\mathcal{D})} \le \sqrt{\frac{\lambda}{\lambda-1}}\, m^{1/\tau-1/2}\, \|f\|_{\mathcal{H}}. \tag{2.6}$$

Proof. Consider $f_m \in \Sigma_m(\mathcal{D})$ and write it as $f_m = \sum_{k \in I_m} a_k g_k$ with $card(I_m) \le m$. By Lemma 2.3, if $m \le 1 + (\lambda M(\mathcal{D}))^{-1}$ then $(\sum_{I_m} |a_k|^2)^{1/2} \le \sqrt{\lambda/(\lambda-1)}\|f\|_{\mathcal{H}}$. By the Bernstein inequality for finite sequences we get

$$\begin{aligned} |f_m|_{\mathcal{K}_\tau^\tau(\mathcal{D})} &\le (\sum_{I_m} |a_k|^\tau)^{1/\tau} \le m^{1/\tau-1/2} (\sum_{I_m} |a_k|^2)^{1/2} \\ &\le \sqrt{\lambda/(\lambda-1)} m^{1/\tau-1/2} \|f\|_{\mathcal{H}}. \end{aligned}$$

□

Proof of Theorem 2.1. Consider $f_m \in \Sigma_m(\mathcal{D})$. Corollary 2.4 proves the Bernstein inequality with exponent $1/\tau - 1/2$ (and *a fortiori* with exponent $\alpha = 2(1/\tau - 1/2)$) and constant $\sqrt{\lambda/(\lambda-1)}$ when $1 \le m \le 1 + (\lambda M(\mathcal{D}))^{-1}$. There remains to get a similar estimate for $m > 1 + (\lambda M(\mathcal{D}))^{-1}$. From Lemma 2.2 we get $|f_m|_{\mathcal{K}_\tau^\tau(\mathcal{D})} \le N^{1/\tau-1/2}\|f_m\|_{\mathcal{H}}$. Since $S(\mathcal{D}) = M(\mathcal{D})\sqrt{N}$, it follows that if $m > 1 + (\lambda M(\mathcal{D}))^{-1} \ge \frac{\sqrt{N}}{\lambda S(\mathcal{D})}$, then

$$|f_m|_{\mathcal{K}_\tau^\tau(\mathcal{D})} \le (\lambda S(\mathcal{D}))^{2/\tau-1}\, m^{2(1/\tau-1/2)} \|f_m\|_{\mathcal{H}}.$$

So for all m, the Bernstein inequality holds with exponent $\alpha = 2(1/\tau - 1/2)$ and constant

$$C(\mathcal{D}) := \min_{\lambda > 1} \max\Big(\sqrt{\frac{\lambda}{\lambda-1}}, (\lambda S(\mathcal{D}))^{2/\tau-1}\Big).$$

Taking $\lambda = 2$ yields the estimate (2.3).

To prove that the exponent $\alpha = 2(1/\tau - 1/2)$ is sharp for $1 \le \tau \le 2$, we will build perfectly incoherent dictionaries $\mathcal{D}$ in $\mathcal{H} := \mathbb{C}^N$ with $N = P^2$ and P an arbitrary large integer P, and exhibit an element $e_0 \in \Sigma_{2P-1}(\mathcal{D}_N)$ with

$$|e_0|_{\mathcal{K}_\tau^\tau(\mathcal{D})} \ge 2^{1/2-1/\tau} \cdot (2P-1)^{2(1/\tau-1/2)} \|f\|_{\mathcal{H}}.$$

For this we let $\mathcal{D}_1 := \{\delta_n\}_{n=0}^{N-1}$ be the Dirac basis for $\mathcal{H}$ and let $\mathcal{D}_2 := \{e_n\}_{n=0}^{N-1}$ be the orthonormal Fourier basis for $\mathcal{H}$. One easily checks that $M(\mathcal{D}_1 \cup \mathcal{D}_2) = 1/\sqrt{N}$. We recall the identity

$$\sum_{k=0}^{P-1} \delta_{k \cdot P} - \sum_{k=0}^{P-1} e_{k \cdot P} = 0, \tag{2.7}$$

which is a consequence of the fact that the "Dirac comb" is invariant under the Fourier transform. We form the dictionary $\mathcal{D} = \mathcal{D}_1 \cup (\mathcal{D}_2 \backslash \{e_0\})$ with $M(\mathcal{D}) = 1/\sqrt{N} = 1/P$. From (2.7) we get

$$e_0 = \sum_{k=0}^{P-1} \delta_{k \cdot P} - \sum_{k=1}^{P-1} e_{k \cdot P},$$

so $e_0 \in \Sigma_{2P-1}(\mathcal{D})$. Now consider an *arbitrary* expansion $e_0 = \sum_{k=0}^{N-1} c_k\delta_k + \sum_{l=1}^{N-1} d_l e_l$ of e_0 in $\mathcal{D}$. By the Hölder inequality we have, with $1 < \tau \le 2$ and $1/\tau + 1/\tau' = 1$,

$$
\begin{aligned}
1 &= |\langle e_0, e_0\rangle| \le \sum_k |c_k||\langle \delta_k, e_0\rangle| \le \Big(\sum_k |c_k|^\tau\Big)^{1/\tau} \cdot \Big(\sum_k |\langle \delta_k, e_0\rangle|^{\tau'}\Big)^{1/\tau'} \\
&\le \Big(\sum_k |c_k|^\tau\Big)^{1/\tau} \cdot N^{1/\tau'} \cdot M(\mathcal{D}) = \Big(\sum_k |c_k|^\tau\Big)^{1/\tau} \cdot N^{1-1/\tau} \cdot N^{-1/2}
\end{aligned}
$$

thus $|e_0|_{\mathcal{K}_\tau^\tau(\mathcal{D})} \ge N^{1/\tau - 1/2} = P^{2(1/\tau-1/2)}$. The same result holds for $\tau = 1$. It follows that we have found $e_0 \in \Sigma_{2P-1}(\mathcal{D})$ which satisfies

$$|e_0|_{\mathcal{K}_\tau^\tau(\mathcal{D})} \ge 2^{-\alpha}(2P-1)^\alpha \|e_0\|_{\mathcal{H}},$$

with $\alpha = 2(1/\tau - 1/2)$. □

3. Decomposable dictionaries

So far, we have obtained Bernstein inequalities for incoherent dictionaries in finite dimensional Hilbert spaces. In Theorem 2.1, what is important is rather the upper estimate (2.3) of the constant in the Bernstein inequality than the inequality itself. Indeed, in finite dimension all norms are equivalent but with equivalence bounds that generally depend on the dimension (see Lemma 2.2). In this section, we will use the results obtained so far to get a Bernstein inequality and a corresponding Bernstein type embedding with **blockwise incoherent dictionaries** in infinite dimensional Hilbert spaces.

In the paper [DH01], Donoho and Huo considered pairs of bases with the so-called *joint block diagonal structure* and the notion of *blockwise mutual coherence* of the bases. As an example of such pairs of bases, they considered in $\mathcal{H} = L^2[0, 2\pi)$ the pair (Meyer-Lemarié wavelets, real bi-sinusoids), see [DH01, Section IX]. We will see more examples in Section 4. Here we are interested in a notion that should generalize the *joint block diagonal structure* of pairs of bases to dictionaries that are not the union of two bases. Considering the notion of **decomposable** dictionary in a general Banach space X, we obtain the second main result of this paper: we reduce the problem of obtaining a Bernstein inequality for decomposable dictionaries to proving "blockwise" Bernstein inequalities with a uniformly bounded constant. Combining this with our first main result from Section 2, we obtain in Section 3.2 a Bernstein inequality for decomposable dictionaries in a Hilbert space that are also uniformly incoherent.

We have the following definition.

Definition 3.1. A dictionary $\mathcal{D}$ in a Banach space X is **decomposable** if one can write $\mathcal{D} = \bigcup_j \mathcal{D}_j$ and decompose X as a direct sum $X = \dot{+}_j X_j$ where for each j, $\mathcal{D}_j$ is a dictionary for the subspace X_j, and the projection P_j onto X_j exists as a bounded linear operator on X. Thus, for any $f \in X$, we have $f = \sum_j P_j f$. We will refer to $\big(\mathcal{D}_j, X_j\big)_j$ as a **decomposition** of $\mathcal{D}$.

We will consider particular decompositions of decomposable dictionaries.

Definition 3.2. A decomposition $\left(\mathcal{D}_j, X_j\right)_j$ of a dictionary $\mathcal{D}$ in a Banach space X is p**-besselian** if there exists $B < \infty$ such that

$$\forall f \in X, \qquad \Big(\sum_j \|P_j f\|_X^p\Big)^{1/p} \leq B\|f\|_X. \tag{3.1}$$

The decomposition is p**-hilbertian** if there exists $A > 0$ such that

$$\forall f \in X, \qquad A\|f\|_X \leq \Big(\sum_j \|P_j f\|_X^p\Big)^{1/p}. \tag{3.2}$$

If $\mathcal{D}$ admits a decomposition that is simultaneously p-besselian and p-hilbertian it is said to be p**-decomposable**.

3.1. **Bernstein inequalities.** In Section 4 we will consider some examples of decomposable dictionaries, but let us immediately state the main result: in any dictionary that admits a p-besselian decomposition, it is necessary and sufficient to prove the Bernstein inequality *blockwise* with a uniform constant.

Theorem 3.3. *Assume $\left(\mathcal{D}_j, X_j\right)_j$ is a p-besselian decomposition of a dictionary $\mathcal{D}$ in X. Let $0 < \tau < p$ and $\alpha \geq 1/\tau - 1/p$. The "global" Bernstein inequality for $\mathcal{K}_\tau^\tau(\mathcal{D})$ with exponent α*

$$|f_m|_{\mathcal{K}_\tau^\tau(\mathcal{D})} \leq Cm^\alpha \|f_m\|_X, \qquad m \geq 1, f_m \in \Sigma_m(\mathcal{D}) \tag{3.3}$$

holds if, and only if, there is a uniform constant C (independent of j) for which the Bernstein inequalities

$$|f_m|_{\mathcal{K}_\tau^\tau(\mathcal{D}_j)} \leq Cm^\alpha \|f_m\|_X, \qquad m \geq 1, \forall j,\ f_m \in \Sigma_m(\mathcal{D}_j), \tag{3.4}$$

hold.

Proof. The fact that (3.3) implies (3.4) follows from the equality $|f|_{\mathcal{K}_\tau^\tau(\mathcal{D}_j)} = |f|_{\mathcal{K}_\tau^\tau(\mathcal{D})}$ for any j and $f \in X_j$. For a general decomposable dictionary, this equality can be proved using the fact that $P_j : X \to X$ is a bounded operator, but we need arguments based on the original topological definition of the sparsity spaces [DT96, GN04a]. For the sake of simplicity, we give a shorter proof assuming that the expression of the sparsity norm (1.1) holds, for $\mathcal{D}$ as well as $\mathcal{D}_j$. As $\mathcal{D}_j \subset \mathcal{D}$ we have $|f|_{\mathcal{K}_\tau^\tau(\mathcal{D})} \leq |f|_{\mathcal{K}_\tau^\tau(\mathcal{D}_j)}$, let us now prove the converse inequality. By definition we can find $\boldsymbol{c}$ with $f = T\boldsymbol{c}$ and $\|\boldsymbol{c}\|_{\ell^\tau} = |f|_{\mathcal{K}_\tau^\tau(\mathcal{D})}$ (remember that T is the synthesis operator, see *e.g.* Definition 1.2). Restricting $\boldsymbol{c}$ to the indices of elements of $\mathcal{D}_j$, we get $\widetilde{\boldsymbol{c}}$ with $\|\widetilde{\boldsymbol{c}}\|_{\ell^\tau} \leq \|\boldsymbol{c}\|_{\ell^\tau}$ and $T\widetilde{\boldsymbol{c}} = P_j T\boldsymbol{c} = P_j f = f$, and obtain the inequality.

We prove that (3.4) implies (3.3) as follows. For $f_m \in \Sigma_m(\mathcal{D})$ we can write $f_m = \sum_j P_j f_m$ (where the sum is finite over $j \in J$ with $card(J) < \infty$), with $P_j f_m \in \Sigma_{m_j}(\mathcal{D}_j)$ and $\sum_j m_j = m$. By the Hölder inequality we get from (3.4), for $s := p/\tau$ and with the usual notation $1/s + 1/s' = 1$

$$\begin{aligned} \sum_j |P_j f_m|_{\mathcal{K}_\tau^\tau(\mathcal{D}_j)}^\tau &\leq C^\tau \sum_j m_j^{\alpha\tau} \|P_j f_m\|_X^\tau \\ &\leq C^\tau \left\|\{m_j^{\alpha\tau}\}_j\right\|_{\ell^{s'}} \left\|\left\{\|P_j f_m\|_X^\tau\right\}_j\right\|_{\ell^s} \end{aligned}$$

As $\alpha \geq 1/\tau - 1/p$, we can check that $\alpha\tau s' \geq 1$, and it follows that $\ell^1 \hookrightarrow \ell^{\alpha\tau s'}$. Hence we have $\left\|\{m_j^{\alpha\tau}\}_j\right\|_{\ell^{s'}} = \|\{m_j\}_j\|_{\ell^{\alpha\tau s'}}^{\alpha\tau} \leq \|\{m_j\}_j\|_{\ell^1}^{\alpha\tau} = m^{\alpha\tau}$ as well as $\left\|\left\{\|P_j f_m\|_X^\tau\right\}_j\right\|_{\ell^s} = \left\|\left\{\|P_j f_m\|_X\right\}_j\right\|_{\ell^{\tau s}}^\tau = \left\|\left\{\|P_j f_m\|_X\right\}_j\right\|_{\ell^p}^\tau \leq B^\tau \|f_m\|_X^\tau$ which follows from the p-besselian property of the decomposition. At this point we have the inequality

$$\sum_j |P_j f_m|_{\mathcal{K}_\tau^\tau(\mathcal{D}_j)}^\tau \leq C^\tau B^\tau m^{\alpha\tau} \|f_m\|_X^\tau,$$

and we conclude using the inequality $|f_m|_{\mathcal{K}_\tau^\tau(\mathcal{D})}^\tau \leq \sum_j |P_j f_m|_{\mathcal{K}_\tau^\tau(\mathcal{D}_j)}^\tau$. □

3.2. **(Highly redundant) blockwise incoherent dictionaries.** The pair of bases (Meyer-Lemarié wavelets, real bi-sinusoids) in $\mathcal{H} = L^2[0, 2\pi)$ is an example of 2-decomposable dictionary with a particular structure : each piece $\mathcal{D}_j$ of the decomposition is an incoherent dictionary in X_j. In this section, we will see that there exists highly redundant similarly *blockwise incoherent* dictionaries, and we will show that it is possible to combine Theorems 2.1 and 3.3 to get a Bernstein embedding even for these highly redundant dictionaries.

Definition 3.4. A dictionary $\mathcal{D}$ in a Hilbert space $\mathcal{H}$ is **blockwise incoherent** if there exists a constant S and a 2-besselian decomposition $\left(\mathcal{D}_j, X_j\right)_j$ of $\mathcal{D}$ where for all j, $\mathcal{D}_j$ *contains an orthonormal basis* $\mathcal{B}_j$ of X_j, $N_j := \dim X_j < \infty$ and $M(\mathcal{D}_j) \leq S/\sqrt{N_j}$. The **separation factor** $S(\mathcal{D})$ is the infimum of the constant S over all admissible decompositions. We say that $\mathcal{D}$ is **blockwise perfectly incoherent** if $S(\mathcal{D}) = 1$.

Blockwise incoherent dictionaries that are simple pairs of bases have a rather low redundancy, and they form tight frames. It is however possible to have *highly redundant* dictionaries that are not frames but are still blockwise perfectly incoherent. We will use the following Theorem to build examples of such dictionaries. We refer to [CCKS97, SH02] for a proof of Theorem 3.5.

Theorem 3.5. *Let* $N_j = 2^{j+1}$, $j \geq 0$ *and consider* $\mathcal{H}_j = \mathbb{R}^{N_j}$. *There exists a tight frame* $\mathcal{D}_j$ *in* $\mathcal{H}_j$ *consisting of the union of* $2^j = N_j/2$ *orthonormal bases for* $\mathcal{H}_j$, *such that for any pair* $u, v \in \mathcal{D}_j, u \neq v$: $|\langle u, v\rangle| \in \{0, N_j^{-1/2}\}$.

For $N_j = 2^j$, $j \geq 0$ *and* $\mathcal{H}_j = \mathbb{C}^{N_j}$, *one can find a tight frame* $\mathcal{D}_j$ *in* $\mathcal{H}_j$ *consisting of the union of* $N_j + 1$ *orthonormal bases for* $\mathcal{H}_j$, *again with the perfect separation property:* $u, v \in \mathcal{D}_j, u \neq v \Rightarrow |\langle u, v\rangle| \in \{0, N_j^{-1/2}\}$.

The frames from Theorem 3.5 are called **Grassmannian frames** due to the fact that their construction is closely related to the Grassmannian packing problem, see [SH02]. In the complex case, $\mathcal{D}_j$ can be a union of $N_j + 1$ maximally mutually incoherent orthonormal bases for $\mathcal{H}_j = \mathbb{C}^{N_j}$, in which case $\mathcal{D}_j$ is of size $N_j(N_j + 1)$ and the frame bound is $N_j + 1$. For the real case, the size of $\mathcal{D}_j$ can be as high as $N_j^2/2$ with a frame bound $N_j/2$. Now it is easy to build a highly redundant blockwise perfectly incoherent dictionary in $\mathcal{H} = \ell^2(\mathbb{N})$.

Example 3.6. We consider the orthogonal direct sum $\ell^2(\mathbb{N}) = \bigoplus_{j=1}^\infty X_j$ with $X_j = \mathbb{R}^{2^{j+1}}$, and take a Grassmannian frame $\mathcal{D}_j$ for X_j given by Theorem 3.5.

By construction, the dictionary $\mathcal{D} = \cup_{j=1}^{\infty} \mathcal{D}_j$ is decomposable and each piece is perfectly incoherent, but we notice that $\mathcal{D}$ is far from being a frame. There cannot be any upper frame bound for $\mathcal{D}$ since the upper frame bound for $\mathcal{D}_j$ is 2^j. Clearly, there is an analog example with $X_j = \mathbb{C}^{2^j}$. $\square$

For blockwise incoherent dictionaries in a Hilbert space $\mathcal{H}$, a Bernstein inequality holds by Theorem 3.3 together with Theorem 2.1, because the upper bound (2.3) on the constant of the Bernstein inequality does not depend on j. A Bernstein type embedding follows using standard results of approximation theory [DL93, Chap. 7].

Theorem 3.7. *Consider $\mathcal{D}$ a blockwise incoherent dictionary in $\mathcal{H}$. Then for $0 < \tau < 2$ and $0 < q \leq \infty$ we have the Bernstein embedding*

$$\mathcal{A}_q^{\gamma}(\mathcal{D}) \hookrightarrow \Big(\mathcal{H}, \mathcal{K}_\tau^\tau(\mathcal{D})\Big)_{1/2,q} \tag{3.5}$$

with $\gamma = 1/\tau - 1/2$.

Proof. The Bernstein inequality

$$|f_m|_{\mathcal{K}_\tau^\tau(\mathcal{D})} \leq C m^{\alpha} \|f_m\|_{\mathcal{H}}, \qquad m \geq 1,\ f_m \in \Sigma_m(\mathcal{D}),$$

with exponent $\alpha = 2(1/\tau - 1/2)$ implies [DL93, Chap. 7] that, for $0 < \gamma < \alpha$ and $0 < q \leq \infty$, there is a continuous embedding

$$\mathcal{A}_q^{\gamma}(\mathcal{D}) \hookrightarrow (\mathcal{H}, \mathcal{K}_\tau^\tau(\mathcal{D}))_{\gamma/\alpha,q} \tag{3.6}$$

where the right hand side is an interpolation space (for details on the real method of interpolation, we refer to [DL93, Chap. 7] and [BS88]). Taking $\gamma = 1/\tau - 1/2 = \alpha/2$ yields the result. $\square$

It should be noted that it is an *open problem* whether we have the "natural" interpolation result $\big(\mathcal{H}, \mathcal{K}_\tau^\tau(\mathcal{D})\big)_{1/2,q} = \mathcal{K}_q^\eta(\mathcal{D})$ with $1/\eta = (1/2 + 1/\tau)/2 = 1/2 + \gamma/2$ for incoherent decomposable dictionaries. If this interpolation result is true, we can identify the Bernstein embeddings $\mathcal{A}_q^\gamma(\mathcal{D}) \hookrightarrow \mathcal{K}_q^\eta(\mathcal{D})$ with the line $\gamma = 2(1/\eta - 1/2)$. One can check that the natural interpolation result holds true for the family of localized frames, see [Grö04] for the definition.

3.3. **Jackson inequalities.** We notice that a general blockwise incoherent dictionary in $\mathcal{H}$ need not be ℓ_1^2-hilbertian. This can, e.g., be verified for the dictionaries from Example 3.6. We cannot apply [GN04a, Th. 3.2] to get a Jackson inequality for such dictionaries. However, there are interesting ℓ_1^2-hilbertian blockwise incoherent dictionaries. For example, a finite union of suitable orthonormal bases for $\mathcal{H}$–such as the dictionary corresponding to one of the bases pairs (Meyer-Lemarié wavelet, real bi-sinusoids) in $\mathcal{H} = L^2[0, 2\pi)$ or (Haar system,Walsh system) in $\mathcal{H} = L^2[0, 1)$– are clearly ℓ_1^2-hilbertian. For such ℓ_1^2-hilbertian dictionaries we do have the Jackson estimate. Let us state this as a Corollary to Theorem 3.7 (and of [GN04a, Th. 3.2]).

Corollary 3.8. *Consider $\mathcal{D}$ a blockwise incoherent dictionary in $\mathcal{H}$. Suppose that $\mathcal{D}$ is also ℓ_1^2-hilbertian. Then for $0 < \tau < 2$ and $0 < q \leq \infty$ we have the continuous embeddings*

$$\mathcal{K}_q^\tau(\mathcal{D}) \hookrightarrow \mathcal{A}_q^\alpha(\mathcal{D}) \hookrightarrow \big(\mathcal{H}, \mathcal{K}_\tau^\tau(\mathcal{D})\big)_{1/2,q} \tag{3.7}$$

with $\alpha = 1/\tau - 1/2$.

3.4. **Comparison of the embedding "lines" for $\mathcal{A}_q^\alpha(\mathcal{D})$.** In the case where we have both a Jackson and a Bernstein embedding for $\mathcal{A}_q^\alpha(\mathcal{D})$, it turns out that there is a "gap" between the embeddings, due to the fact that the exponents $\alpha = 1/\tau - 1/2$ in the Jackson inequality and $\alpha = 2(1/\tau - 1/2)$ in the Bernstein inequality do not match up.

The Jackson inequality correspond to the line $\alpha = 1/\tau - 1/2$. This means precisely that if f "has sparsity" τ, then is can be approximated with the rate $\alpha = 1/\tau - 1/2$. Let us prove that, for a p-decomposable dictionary, the Jackson embedding line $\alpha = 1/\tau - 1/p$ is the best possible "point by point".

Theorem 3.9. *Let $\mathcal{D}$ be a p-decomposable dictionary in X. Assume the Jackson type embedding $\mathcal{K}_q^\tau(\mathcal{D}) \hookrightarrow \mathcal{A}_\infty^\alpha(\mathcal{D})$ holds for some $0 < \tau < p$, some $0 < q \leq \infty$ and some $\alpha > 0$. Then $\alpha \leq 1/\tau - 1/p$.*

Proof. Fix $g_j \in \mathcal{D}_j$ a sequence of dictionary vectors. For any $\boldsymbol{c} = \{c_j\} \in \ell^p$ we can consider $f(\boldsymbol{c}) := \sum_j c_j g_j$: by the p-hilbertian property of the decomposition, the series is seen to be unconditionally convergent in X.

The result will follow from the inequality $\sigma_m(\boldsymbol{c}, \mathcal{B})_{\ell^p} \leq B\sigma_m\big(f(\boldsymbol{c}), \mathcal{D}\big)_X$ with $\mathcal{B}$ the canonical basis in ℓ^p. To prove this inequality, we proceed first as in the proof that (3.4) implies (3.3). For $f_m \in \Sigma_m(\mathcal{D})$ we can write (as a finite sum over $j \in J$ with $card(J) \leq m$) $f_m = \sum_j P_j f_m$, with $P_j f_m \in \Sigma_{m_j}(\mathcal{D}_j)$, $\sum_j m_j = m$. By the p-besselian property of the decomposition, it follows that

$$B^p \|f(\boldsymbol{d}) - f_m\|_X^p \geq \sum_j \|c_j g_j - P_j f_m\|_X^p \geq \sum_{j \notin J} |c_j|^p \geq \sigma_m\big(\boldsymbol{c}, \mathcal{B}\big)_{\ell^p}^p.$$

Taking the infimum over $f_m \in \Sigma_m(\mathcal{D})$ we get $\sigma_m(\boldsymbol{c}, \mathcal{B})_{\ell^p} \leq B\sigma_m\big(f(\boldsymbol{c}), \mathcal{D}\big)$.

Let us now prove the Theorem. By assumption we have for some $C < \infty$ and all $f \in X$: $\sigma_m\big(f, \mathcal{D}\big)_X \leq |f|_{\mathcal{A}_\infty^\alpha(\mathcal{D})} m^{-\alpha} \leq C|f|_{\mathcal{K}_q^\tau(\mathcal{D})} m^{-\alpha}$ for $m \geq 1$. Specializing to $f = f(\boldsymbol{c})$ we get for all $\boldsymbol{c} \in \ell^p$ and $m \geq 1$:

$$\sigma_m\big(\boldsymbol{c}, \mathcal{B}\big)_{\ell^p} \leq B\sigma_m\big(f(\boldsymbol{c}), \mathcal{D}\big)_X \leq BC|f(\boldsymbol{c})|_{\mathcal{K}_q^\tau(\mathcal{D})} m^{-\alpha} \leq BC\|\boldsymbol{c}\|_{\ell_q^\tau} m^{-\alpha}.$$

It follows that $\ell_q^\tau \hookrightarrow \mathcal{A}_\infty^\alpha(\mathcal{B}, \ell^p) = \ell_\infty^\eta$ with $1/\eta = \alpha + 1/p$, and this is known to be possible only if $\tau \leq \eta$, *i.e.* if $\alpha \leq 1/\tau - 1/p$. □

The Bernstein inequality for blockwise incoherent dictionaries "essentially" corresponds to the line $\alpha = 2(1/\tau - 1/2)$, and Theorem 2.1 shows that the exponent $\alpha = 2(1/\tau - 1/2)$ cannot be generally improved. So, it is just in the nature of the problem that the lines do generally have different slopes, and $\mathcal{A}_q^\alpha(\mathcal{D})$ just *cannot* be exactly characterized in terms of sparsity spaces for arbitrary blockwise incoherent dictionaries.

4. Examples

We have already seen that examples of blockwise incoherent dictionaries can be built in an abstract framework by concatenating pairs of perfectly incoherent bases in finite dimensional Hilbert spaces. In particular, many examples can be built based on sub-dictionaries of Grassmannian frames (see Example 3.6). In this section, however, we are interested in more "concrete" examples where the dictionaries are built from classical systems used in harmonic analysis.

First we consider the example of the (Haar-Walsh) dictionary on $L^2[0,1)$. In order to enlighten the (partial) characterization of $\mathcal{A}^\alpha_\tau(\mathcal{D})$ given by Corollary 3.8, we try to give some characterization of the sparsity spaces $\mathcal{K}^\tau_\tau(\mathcal{D})$ in terms of classical smoothness spaces. With this aim, we consider the (Dirichlet wavelet,trigonometric system) and show that the sparsity spaces are related to the Wiener algebra and Besov spaces on the interval. We conclude the section by some examples of blockwise incoherent dictionaries on $L^2(\mathbb{R})$ that we obtain by "patching" dictionaries on $L^2[0,1)$.

4.1. **Haar-Walsh dictionary on $L^2[0,1)$.** The prime example of blockwise perfectly incoherent dictionary in $L^2[0,1)$ is given by the (Haar,Walsh) pair.

Example 4.1. Let $\mathcal{B}_H = \{h_n\}_{n=0}^\infty$ be the Haar system on $[0,1)$ with the standard ordering, and let $\mathcal{B}_W = \{W_n\}_{n=0}^\infty$ be the Walsh system on $[0,1)$ with the Paley ordering, see *e.g.* [GES91]. Then $\mathcal{D} = \mathcal{B}_H \cup \mathcal{B}_W$ is a 2-decomposable dictionary in $X = L^2[0,1)$ with $X_0 = \text{span}\{1\}$ and for $j \geq 1$, $X_j = \text{span}\{h_n\}_{n=2^{j-1}}^{2^j-1} = \text{span}\{W_n\}_{n=2^{j-1}}^{2^j-1}$. with $\mathcal{D}_j = \mathcal{B}_{H,j} \cup \mathcal{B}_{W,j}$, where $\mathcal{B}_{H,j}$ is the set of Haar functions at scale j, $\mathcal{B}_{W,j}$ the set of Walsh functions at "frequency" indices $[2^{j-1}, 2^j - 1]$. We have $N_0 = 1$ and $N_j = \dim(X_j) = 2^{j-1}, j \geq 1$. One can check, see *e.g.* [GES91], $|\langle h, w\rangle| = N_j^{-1/2}$ for any Haar wavelet h and Walsh function w in the space X_j, *i.e.* $\mathcal{B}_{H,j}$ and $\mathcal{B}_{W,j}$ are maximally mutually incoherent.

Corollary 4.2. *For $\mathcal{D} = \mathcal{B}_H \cup \mathcal{B}_W$ the union of the Haar and Walsh orthonormal bases in $\mathcal{H} = L^2[0,1]$, and $0 < \tau < 2$ the embeddings* (3.7) *hold true.*

4.2. **Dirichlet wavelets and trigonometric system.** The Haar and Walsh systems both consist of only piecewise continuous functions, which may be a problem when analyzing very smooth functions. A smooth analogue is given by the (Meyer-Lemarié wavelets, real bi-sinusoids) example of Donoho and Huo [DH01], but let us consider here a similar example, given by the trigonometric system and the "Dirichlet wavelet" [Woj97, Tem99, Tem00]. Define $e_m : t \mapsto e^{2i\pi mt}$, $m \in \mathbb{Z}$. For $j \geq 0$ and $k = 0, 1, \ldots, 2^j - 1$, we define

$$\Psi_{j,k}(t) := \Psi_{j,0}(t - k2^{-j}) \tag{4.1}$$

with

$$\Psi_{j,0}(t) := 2^{-j/2} \sum_{m=2^j}^{2^{j+1}-1} e_m(t) \tag{4.2}$$

The system $\{e_0\} \cup \{\Psi_{j,k},\ j \geq 0,\ k = 0, 1, \ldots, 2^j - 1\}$ is an orthonormal basis for $H^2(0,1)$, see [Woj97], and can easily be extended to an orthonormal basis for $L^2(0,1)$ using $\{\overline{\Psi_{j,k}(x)},\ j \geq 0,\ k = 0, 1, \ldots, 2^j - 1\}$.

Example 4.3. Consider $\mathcal{B}_T = \{e_m\}_{m\geq 0}$ and $\mathcal{B}_D = \{e_0\} \cup \{\Psi_{j,k}\}_{j\geq 0,k}$, the trigonometric and "Dirichlet wavelet" orthonormal bases for $H^2(0,1)$, respectively. Let $\mathcal{D} = \mathcal{B}_T \cup \mathcal{B}_D$: it is easily seen to be 2-decomposable in $H^2(0,1) = \oplus_j X_j$, where

$X_{-1} = \mathrm{span}(e_0)$, and $X_j = \mathrm{span}(e_m)_{m=2^j}^{2^{j+1}-1} = \mathrm{span}(\Psi_{j,k})_{k=0}^{2^j-1}$, $j \geq 0$. We have $N_j = \dim(X_j) = 2^j$, $j \geq 0$. As above one easily checks that the dictionary $\mathcal{D}_j$ in X_j is perfectly incoherent. The reader can check that the result has a straightforward extension to $L^2(0,1)$.

As $\mathcal{D} = \mathcal{B}_T \cup \mathcal{B}_D$ forms an incoherent dictionary in $H^2(0,1)$ (resp. in $L^2(0,1)$) consisting of smooth atoms we have the Corollary:

Corollary 4.4. *For $\mathcal{D} = \mathcal{B}_T \cup \mathcal{B}_D$ the union of the trigonometric and "Dirichlet wavelet" orthonormal bases in $\mathcal{H} = H^2(0,1)$ (resp. in $L^2(0,1)$), and $0 < \tau < 2$ the embeddings* (3.7) *hold true.*

Next we will (partially) characterize the sparsity spaces $\mathcal{K}_q^\tau(\mathcal{B}_T \cup \mathcal{B}_D)$ in terms of classical smoothness spaces. First, let us verify that it is possible to completely characterize the Besov space $B_q^s(L^p[0,1))$ using the Dirichlet wavelet. We need the following Lemma, which we have included for the sake of completeness.

Lemma 4.5. *Let P_j be the projection defined on $L^2[0,1)$ for $j \geq 0$ by*

$$P_j\left(\sum_{m\in\mathbb{Z}} a_m e_m(t)\right) = \sum_{2^j \leq |m| < 2^{j+1}} a_m e_m(t).$$

Then, for $1 < p < \infty$, $0 < q \leq \infty$ and $s > 0$

$$|f|_{B_q^s(L^p[0,1))} \asymp \left(\sum_{j=0}^{\infty} [2^{js}\|P_j f\|_p]^q\right)^{1/p}$$

where the constants in the equivalence may depend on p, q, s.

Proof. Let $\phi \in C^\infty(\mathbb{R})$ with $\mathrm{supp}(\phi) \subset \{\xi : \pi/2 < |\xi| < 2\pi\}$ such that $\sum_{j\in\mathbb{Z}} \phi(2^{-j}\xi) = 1$ for $\xi \neq 0$. Define $\Phi_j(t) = \sum_{m\in\mathbb{Z}} \phi(2\pi m 2^{-j}) e_m(t)$. Then by definition (see [Tri83])

$$|f|_{B_q^s(L^p[0,1))} := \left(\sum_{j=1}^{\infty} [2^{sj}\|\Phi_j * f\|_p]^q\right)^{1/p}.$$

Notice that $\|\Phi_1 * f\|_p = \|P_0 f\|_p$ and using the support restriction imposed on ϕ and the fact that the P_j's are uniformly bounded on $L^p[0,1)$ by the Riesz projection theorem, we have for $j \geq 1$,

$$\|P_j f\|_p = \|P_j[(\Phi_{j+1} + \Phi_{j+2}) * f]\|_p \leq C\|(\Phi_{j+1} + \Phi_{j+2}) * f\|_p,$$

and conversely for $j \geq 0$,

$$\|\Phi_{j+2} * f\|_p = \|\Phi_{j+2} * [(P_j f + P_{j+1})f]\|_p \leq C\|(P_j f + P_{j+1})f\|_p,$$

where uniform boundedness of the convolution operators induced by the Φ_j's follows by using the Hörmander-Mihlin multiplier theorem. Using these two estimates we reach the wanted conclusion. ∎

Using Lemma 4.5 we can easily get the following characterization.

Corollary 4.6. *For the Dirichlet wavelet system defined by* (4.1) *and* (4.2)*, which is normalized in* $L^2(0,1)$*, and for any* $1 < p < \infty$*,*

$$|f|_{B^s_q(L^p[0,1))} \asymp \left(\sum_{j=0}^{\infty} 2^{j(s+1/2-1/p)q} \left(\sum_{k=0}^{2^j-1} \left(|\langle f, \Psi_{j,k}\rangle|^p + |\langle f, \overline{\Psi_{j,k}}\rangle|^p \right) \right)^{q/p} \right)^{1/q}.$$

Proof. A consequence of Wojtaszczyk's result [Woj97] is the fact that, uniformly in j,

$$\|P_j \tilde{f}\|_p \asymp 2^{j(1/2-1/p)} \left(\sum_{k=0}^{2^j-1} |\langle \tilde{f}, \Psi_{j,k}\rangle|^p \right)^{1/p},$$

where $\tilde{f}$ denotes the orthogonal projection onto $\mathrm{span}\{e_k\}_{k\geq 0}$ of f. From the isomorphism between $H^p(0,1)$ and $L^p(0,1)$, see [Woj97] we have $\|P_j f\|_p \asymp (\|P_j \tilde{f}\|_p^p + \|P_j \tilde{\bar{f}}\|_p^p)^{1/p}$ and the conclusion follows using Lemma 4.5. □

Remark 4.7. For $1 < p < \infty$, as $\|\Psi_{j,k}\|_p \asymp 2^{j(1/2-1/p)}$, Corollary 4.6 shows that for the basis $\mathcal{B}^p_D$ of Dirichlet wavelets normalized in L^p we have

$$|f|_{B^\alpha_\tau(L^\tau(0,1))} \asymp |\cdot|_{\mathcal{K}^\tau_\tau(\mathcal{B}^p_D)}, \quad \alpha = 1/\tau - 1/p. \quad \square$$

As the trigonometric system $\mathcal{B}_T$ is simultaneously normalized in $L^p(0,1)$ for all p, we know [GN04c, Cor. 1] that, with $\alpha = 1/\tau - 1/p$,

$$\begin{aligned} \mathcal{K}^\tau_\tau(\mathcal{B}_T \cup \mathcal{B}^p_D, L^p[0,1)) &= \mathcal{K}^\tau_\tau(\mathcal{B}_T, L^p[0,1)) + \mathcal{K}^\tau_\tau(\mathcal{B}^p_D, L^p[0,1)) \\ &= \mathcal{K}^\tau_\tau(\mathcal{B}_T, L^p[0,1)) + \mathcal{B}^\alpha_\tau(L^\tau[0,1)). \end{aligned}$$

The sparsity space $\mathcal{K}^\tau_\tau(\mathcal{B}_T, L^p[0,1))$ is exactly the range of the Fourier transform on $\ell^\tau(\mathbb{Z})$, so a prominent special case is $\mathcal{K}^1_1(\mathcal{B}_T, L^p[0,1)) = W(\mathbb{Z})$, the *Wiener algebra.*

In Corollary 4.4, we have to consider the case $p = 2$, thus we get two sided embeddings between the approximation classes and (interpolation of) sparsity spaces. On the Jackson side, we have to consider

$$\mathcal{K}^\tau_\tau(\mathcal{B}_T \cup \mathcal{B}_D) = \mathcal{K}^\tau_\tau(\mathcal{B}_T) + \mathcal{B}^\alpha_\tau(L^\tau[0,1))$$

with $\alpha = 1/\tau - 1/2$. The norm in such a space is characterized by

$$\begin{aligned} |f|_{\mathcal{K}^\tau_\tau(\mathcal{B}_T \cup \mathcal{B}_D)} &= \inf_{g \in \mathcal{B}^\alpha_\tau(L^\tau[0,1))} |f-g|_{\mathcal{K}^\tau_\tau(\mathcal{B}_T)} + |g|_{\mathcal{B}^\alpha_\tau(L^\tau[0,1))} \\ &= \min_{g \in \mathcal{B}^\alpha_\tau(L^\tau[0,1))} |f-g|_{\mathcal{K}^\tau_\tau(\mathcal{B}_T)} + |g|_{\mathcal{B}^\alpha_\tau(L^\tau[0,1))}. \end{aligned}$$

Although the norm on $\mathcal{K}^\tau_\tau(\mathcal{B}_T \cup \mathcal{B}_D)$ is given explicitly, one would like a more *straightforward* alternative expression for this norm. However, more information about $\mathcal{K}^\tau_\tau(\mathcal{B}_T)$ and $\mathcal{B}^\alpha_\tau(L^\tau(0,1)) \cap \mathcal{K}^\tau_\tau(\mathcal{B}_T)$ is needed in order to derive such a characterization, and we leave this as a challenge for the reader. Another open question is how to characterize the interpolation spaces $\left(L^2[0,1), \mathcal{K}^\tau_\tau(\mathcal{B}_T \cup \mathcal{B}^p_D, L^p[0,1))\right)_{\theta,q}$. It is unlikely that this interpolation family can be expressed in terms of any classical function spaces.

4.3. **Examples on $L^2(\mathbb{R})$: patching dictionaries for $L^2[0,1)$.** Based on examples of blockwise incoherent dictionaries in $L^2[0,1)$, there is a simple construction on $L^2(\mathbb{R})$. Indeed, for any $1 < p < \infty$, we can construct a p-decomposable dictionary in $X = L^p(\mathbb{R})$ as follows.

Example 4.8. Let P_n be the projection given by $P_n f(x) = \chi_{[n,n+1)}(x)f(x)$. Clearly $L^p(\mathbb{R}) = \dot{+}_n P_n L^p(\mathbb{R})$,

$$\|f\|_p = \left(\sum_{n\in\mathbb{Z}} \|P_n f\|_p^p\right)^{1/p},$$

and as dictionary $\mathcal{D}_n$ in $P_n L^2(\mathbb{R})$ we can take any (perfectly) incoherent dictionary such as the (Haar,Walsh) system or the (Dirichlet wavelet, trigonometric) system on $[0,1)$ translated to $[n,n+1)$. The resulting dictionary $\mathcal{D} = \cup_n \mathcal{D}_n$ is blockwise (perfectly) incoherent.

This example can easily be refined by using smooth projections associated with a uniform partition of $\mathbb{R}$, see [AWW92]. Dictionaries associated with such a smooth partition can then be created using Wickerhauser's method to construct smooth localized bases [Wic93].

5. Conclusion

We have studied Bernstein inequalities for some classes of redundant dictionaries in a Banach space. For dictionaries in a Hilbert space, that are blockwise incoherent, we have derived an explicit sharp Bernstein inequality. Using examples based on Grassmannian frames we have shown that the Bernstein inequality is valid even for some highly redundant dictionaries. For 2-hilbertian blockwise incoherent dictionaries in a Hilbert space, we have both a sharp Bernstein and a sharp Jackson inequality. The exponents of the inequalities do not match so one cannot quite get a complete characterization of the associated approximation spaces, but only a two-sided embedding of the approximation spaces. It is in the nature of the approximation space that it cannot be exactly characterized in terms of sparsity spaces in this case, and by combining two mutually incoherent bases, one gets an approximation space that is strictly larger than the sum of the individual approximation spaces.

Acknowledgments

The authors would like to thank Ronald A. DeVore for introducing us to most of the problems discussed in this paper. The idea of considering nonlinear approximation with incoherent dictionaries emerged as a follow-up to David Donoho's interesting talk (based on [DH01] during the second workshop of the *Wavelet Ideal Data Representation Center*, at the Shannon Laboratory of AT&T Labs, New Jersey, in October 1999.

References

[AST01] A. Aldroubi, Q. Sun, and W.-S. Tang. p-frames and shift invariant subspaces of l^p. *J. Fourier Anal. Appl.*, 7(1):1–21, 2001.

[AWW92] P. Auscher, G. Weiss, and M. V. Wickerhauser. *Wavelets*, chapter Local sine and cosine bases of Coifman and Meyer and the construction of smooth wavelets., pages 237–256. Academic Press, Boston, MA, 1992.

[BS88] Colin Bennett and Robert Sharpley. *Interpolation of operators*. Academic Press Inc., Boston, MA, 1988.

[CCKS97] A. R. Calderbank, P. J. Cameron, W. M. Kantor, and J. J. Seidel. Z_4-Kerdock codes, orthogonal spreads, and extremal Euclidean line-sets. *Proc. London Math. Soc. (3)*, 75(2):436–480, 1997.

[CD99] S. Chen and D.L. Donoho. Atomic decomposition by basis pursuit. *SIAM Journal on Scientific Computing*, 20(1):33–61, January 1999.

[Chr02] Ole Christensen. *An introduction to frames and Riesz bases*. 2002. Manuscript.

[DH01] D.L. Donoho and Xiaoming Huo. Uncertainty principles and ideal atomic decompositions. *IEEE Trans. Inform. Theory*, 47(7):2845–2862, November 2001.

[DE03] D. L. Donoho and M. Elad. Optimally sparse representation in general (nonorthogonal) dictionaries via l^1 minimization. *Proc. Natl. Acad. Sci. USA*, 100(5):2197–2202 (electronic), 2003.

[DHRS01] Ingrid Daubechies, Bin Han, Amos Ron, and Zuowei Shen. Framelets: MRA-based constructions of wavelet frames. *Preprint*, 2001.

[DKKT01] S.J. Dilworth, N.J. Kalton, D. Kutzarova, and V.N. Temlyakov. The thresholding greedy algorithm, greedy bases, and duality. Technical Report 0123, Dept of Mathematics, University of South Carolina, Columbia, SC 29208, 2001.

[DL93] Ronald A. DeVore and George G. Lorentz. *Constructive approximation*. Springer-Verlag, Berlin, 1993.

[DT96] R. A. DeVore and V. N. Temlyakov. Some remarks on greedy algorithms. *Adv. Comput. Math.*, 5(2-3):173–187, 1996.

[GES91] B. Golubov, A. Efimov, and V. Skvortsov. *Walsh series and transforms*. Theory and applications, Translated from the 1987 Russian original by W. R. Wade. Kluwer Academic Publishers Group, Dordrecht, 1991.

[GN01] R. Gribonval and M. Nielsen. Some remarks on nonlinear approximation with Schauder bases. *East J. Approx.*, 7(3):267–285, 2001.

[GN03] R. Gribonval and M. Nielsen. Sparse representations in unions of bases. *IEEE Trans. Inform. Theory*, 49(12):3320–3325, 2003.

[GN04a] R. Gribonval and M. Nielsen. Nonlinear approximation with dictionaries. I. Direct estimates. *J. Fourier Anal. Appl.*, 10(1):51-71, 2004.

[GN04b] R. Gribonval and M. Nielsen. On a problem of Gröchenig about nonlinear approximation with localized frames. *J. Fourier Anal. Appl.*, 10(4):433-437, 2004.

[GN04c] R. Gribonval and M. Nielsen. On approximation with spline generated framelets. *Constr. Approx.*, 20(2):207-232, 2004.

[Grö00] Karlheinz Gröchenig. *Foundations of Time-Frequency Analysis (Applied Numerical and Harmonic Analysis)*. Birkhauser, December 2000. ISBN: 0817640223.

[GS00] K. Gröchenig and S. Samarah. Nonlinear approximation with local Fourier bases. *Constr. Approx.*, 16(3):317–332, 2000.

[Grö03] K. Gröchenig. Localized frames are finite unions of Riesz sequences. *Adv. Comput. Math.*, 18(2-4):149–157, 2003.

[Grö04] K. Gröchenig. Localization of frames, Banach frames, and the invertibility of the frame operator. *J. Fourier Anal. Appl.*, 10(2):105–132, 2004.

[KP01] G. Kerkyacharian and D. Picard. Entropy, universal coding, approximation and bases properties. Technical Report Preprint No. 663, Universities Paris 6 & 7, 2001.

[Pet88] Pencho P. Petrushev. Direct and converse theorems for spline and rational approximation and Besov spaces. In *Function spaces and applications (Lund, 1986)*, pages 363–377. Springer, Berlin, 1988.

[SBT00] S. Sardy, A. Bruce, and P. Tseng. Block coordinate relaxation methods for nonparametric wavelet denoising. *Journal of Computational and Graphical Statistics*, 9(2), 2000.

[SH02] T. Strohmer and R. Heath. Grassmannian frames with applications to coding and communications. Technical report, 2002. Preprint, submitted to Appl.Comp.Harm.Anal.

[Ste55] S. B. Stechkin. On absolute convergence of orthogonal series. *Dok. Akad. Nauk SSSR*, 102:37–40, 1955.

[Tem99] V.N. Temlyakov. Universal bases and greedy algorithms. Technical Report 9908, Dept of Mathematics, University of South Carolina, Columbia, SC 29208, 1999.

[Tem00] V.N. Temlyakov. Greedy algorithms with regard to multivariate systems with special structure. *Constr. Approx.*, 16(3):399–425, 2000.

[Tri83] Hans Triebel. *Theory of function spaces*, volume 78 of *Monographs in Mathematics*. Birkhäuser Verlag, Basel, 1983.

[Wic93] M.V. Wickerhauser. Smooth localized orthonormal bases. *C. R. Acad. Sci. Paris Sér. I Math.*, 316(5):423–427, 1993.

[Woj97] P. Wojtaszczyk. On unconditional polynomial bases in L_p and Bergman spaces. *Constr. Approx.*, 13(1):1–15, 1997.

IRISA-INRIA, Campus de Beaulieu, 35042 Rennes cedex, France
E-mail address: remi.gribonval@inria.fr

Department of Mathematical Sciences, Aalborg University, Fredrik Bajers Vej 7G, DK-9220 Aalborg East, Denmark
E-mail address: mnielsen@math.aau.dk